

# A Benchmark for Evaluation of Algorithms for Identification of Cellular Correlates of Clinical Outcomes

Nima Aghaeepour<sup>\*,1,2,3</sup>, Pratip Chattopadhyay<sup>\*,4</sup>, Maria Chikina<sup>5</sup>, Tom Dhaene<sup>6</sup>, Sofie Van Gassen<sup>6,7,8</sup>, Miron Kursa<sup>9</sup>, Bart N. Lambrecht<sup>7,8</sup>, Mehrnoush Malek<sup>1</sup>, Yu Qian<sup>10</sup>, Peng Qiu<sup>11</sup>, Yvan Saeys<sup>7,8</sup>, Rick Stanton<sup>10</sup>, Dong Tong<sup>12</sup>, Celine Vens<sup>7,8,13</sup>, Sławomir Walkowiak<sup>9</sup>, Kui Wang<sup>14,15</sup>, Greg Finak<sup>\*,16</sup>, Raphael Gottardo<sup>\*,16</sup>, Tim Mosmann<sup>\*,17</sup>, Garry Nolan<sup>\*,3</sup>, Richard H. Scheuermann<sup>\*,10,18</sup>, Ryan R. Brinkman<sup>\*,+,1,2</sup>

<sup>1</sup> British Columbia Cancer Agency, Vancouver, BC, Canada.

<sup>2</sup> Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada.

<sup>3</sup> Baxter Laboratory in Stem Cell Biology, Stanford University, Stanford, CA, USA.

<sup>4</sup> ImmunoTechnology Section, Vaccine Research Center, National Institutes of Health, Washington, DC, USA

<sup>5</sup> Department Computational and Systems Biology, University of Pittsburgh, Pittsburg, USA

<sup>6</sup> Department of Information Technology, Ghent University - iMinds, Ghent, Belgium

<sup>7</sup> Inflammation Research Center, VIB, Ghent, Belgium

<sup>8</sup> Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium

<sup>9</sup> Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warsaw, Poland

<sup>10</sup> J. Craig Venter Institute, La Jolla, CA, USA.

<sup>11</sup> Department of Biomedical Engineering, Georgia Institute of Technology and Emory University

<sup>12</sup> The John van Geest Cancer Research Centre, Nottingham Trent University, UK & Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

<sup>13</sup> Department of Public Health and Primary Care, KU Leuven Kulak, Kortrijk, Belgium

<sup>14</sup> Department of Mathematics, University of Queensland, St. Lucia, Brisbane, Australia

<sup>15</sup> School of Medicine, Shihezi University, Xinjiang 832000, China

<sup>16</sup> Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

<sup>17</sup> David H. Smith Center for Vaccine Biology and Immunology, University of Rochester Medical Center, Rochester, NY, USA.

<sup>18</sup> Department of Pathology, University of California, San Diego, CA, USA

\*Equal Contribution

+Corresponding Author

## Abstract

The Flow Cytometry: Critical Assessment of Population Identification Methods (FlowCAP) challenges were established to compare the performance of computational methods for identifying cell populations in multidimensional flow cytometry data. Here we report the results of FlowCAP-IV where algorithms from seven different research groups predicted the time to

progression to AIDS among a cohort of 384 HIV+ subjects, using antigen-stimulated peripheral blood mononuclear cell (PBMC) samples analyzed with a 14-color staining panel. Two approaches (FlowReMi.1 and flowDensity-flowType-RchyOptimyx) provided statistically significant predictive value in the blinded test set. Manual validation of submitted results indicated that unbiased analysis of single cell phenotypes could reveal unexpected cell types that correlated with outcomes of interest in high dimensional flow cytometry datasets.

## Keywords

flow cytometry, bioinformatics, clustering, classification, data analysis, HIV, clinical outcome, supervised analysis

## Introduction

A wide range of computational tools is required for systematic analysis of high-dimensional and high-throughput single cell data in different clinical and experimental settings. Despite recent developments in flow cytometry bioinformatics [1], guidance for end users regarding the appropriate use of these methods has been scarce. The FlowCAP Consortium has developed a series of benchmarks for objective evaluation of computational tools in different settings. FlowCAP-I focused on the evaluation of cell population identification algorithms compared to manual analysis by human experts for a selection of different cell populations across several datasets [2]. While this approach was helpful to demonstrate that algorithms can identify known populations, it did not directly address the effectiveness of algorithms that identified cell types not previously delineated by the human analysts. To address this limitation the next FlowCAP challenges focused on the identification of cellular correlates of an independent biological variable (e.g. a clinical outcome). While these challenges demonstrated that computational analysis could outperform manual analysis for identification of cell types correlated with external factors, they did not establish a benchmark for evaluating the relative performance of data analysis pipelines because in one challenge many algorithms were able to achieve perfect classification accuracy, whereas in another challenge none of the algorithms (or human expert) could successfully identify a cell type that correlated with the clinical outcome. Therefore, a more challenging benchmark for evaluation of these algorithms was required to help evaluate their relative performance. In this article we report the results of the latest FlowCAP challenge that addresses this requirement through the use of a complex dataset to evaluate data analysis strategies for identification of correlates of clinical outcome – in this case the time to progression to AIDS in HIV positive participants.

## Materials and Methods

We used data from a long-standing natural history study of HIV [3], from which 384 individuals were selected, with known HIV seroconversion dates (so that an estimated date of infection could be calculated) and long-term clinical follow-up [4]. For each individual, the time to AIDS diagnosis (“survival time”) was known, and patients were censored upon initiation of treatment or when lost to follow-up. We chose the earliest available cell specimen (<18 months post-infection (p.i.); mean = 7 months p.i.); our previous analysis showed that the timing of the cell specimen had no relationship to survival time. Cells were stained with two polychromatic flow cytometry panels: the first assessed the resting phenotypes of T-cells directly *ex vivo* [4] (but was not used for this study), while the second characterized cytokine and other functional responses to HIV, after stimulation with pools of overlapping peptides from HIV-Gag proteins (Chattopadhyay et al., under review). Specifically, the latter panel examined the cytokines IFN $\gamma$ , IL2, and TNF, along with a costimulatory molecule (CD154) upregulated by CD4+ T-cells upon activation and a marker of T-cell degranulation (CD107a). Cell surface markers that define the maturity of the HIV-specific T-cell were examined in this panel as well (CD45RO, CCR7, CD27, and CD57) so that the functional responses could be ascribed to cell types well-described in the literature (e.g., central memory T-cells, senescent T-cells, etc.). Manual analysis of the dataset was performed previously, and showed that no specific characteristics of the anti-HIV T-cell response in the selected samples could predict survival time (specifically, the following characteristics were tested: total magnitude of the HIV Gag or Env specific T-cell responses, magnitude of individual functional responses, polyfunctionality of the CD4 or CD8 response, and differentiation state of the response; these are described in greater detail in a manuscript under review (Chattopadhyay, et al)). It is important to note, however, that the manual data analysis examined only cell characteristics that were known to the biologists a priori. Manual data analysis of all possible combinations of markers in the dataset was not feasible. All selected samples from divided into two equal training- and test-sets randomly and uniformly. No correlation with the clinical outcome was observed across the two sets (pvalue > 0.22). The FlowCAP participants were blinded to the clinical outcomes in the testset, which was used for independently verifying the submitted results.

To ensure the feasibility of the study, members of the FlowCAP organizing committee not participating in the submission of algorithm results performed an independent analysis [5]. The raw Gag-stimulated samples were pre-processed using an optimized logicle transformation [6] and 1000 cells per patient were selected randomly. K-means clustering was used to identify 50 cell types simultaneously in all patients using the pooled cells. For each cluster, the frequency of the cells (i.e., the number of cells in the cluster divided by the total number of cells) was used for the correlative analysis (Cox-proportional hazards model) on the training set. The median expression of each marker for each cell type as well as log-rank p-values for each cluster were calculated (Supplemental Figure 1). The most significant clusters included CD3+CD4+ and CD3+CD8+ populations that also express a range of other markers (e.g., CD27, CCR7, and IL2) and lacked expression of others (e.g., CD57 and TNF). These clusters remained correlated with the clinical outcome in the test set, suggesting that successful identification of correlates of disease progression in this dataset was possible (Supplemental Figure 2).

Data from unstimulated and Gag-stimulated samples were distributed to all participants. The dataset was randomly partitioned into a training set and a test set of equal sizes. The clinical information (survival time and censorship status) of the patients in the test set was not provided to the participants. Each participant provided a vector of final predictions (extracted from one or a combination of several cell types) that was most correlated with the clinical outcome (as measured by a log-rank test on a Cox proportional hazards regression). The complete source code as well as the required software packages for independent reproduction of the results was also required. The results were evaluated by an independent log-rank test on the blinded test set.

## Results

Twenty-four groups requested to participate in FlowCAP-IV and received the data. We received 9 submissions from 7 groups (Table 1) using combinations of different algorithms including [7]-[17]. Log-rank test p-values for the submitted result indicated that while several algorithms identified cell types strongly correlated with the clinical outcome in the training dataset, only the flowDensity-flowType-RchyOptimyx and FloReMi.1 approaches maintained a significant correlation ( $p < 0.05$ ) in the blinded test set. flowDensity is a supervised sequential bivariate clustering algorithm that mimics manual gating methods. flowType-RchyOptimyx is an approach that uses partitioning of cells into categories (e.g., positive or negative) using 1-D density estimate to calculate cut-offs for each marker and enumerate all cell types in a sample, followed by dynamic programming to efficiently construct k-shortest paths to important cell populations. FloReMi.1 combines flowType for cell population identification with a random forest approach using cell population-based features to build a survival regression model. One version of the three different versions of FloReMi submitted (FloReMi.1) remained significantly correlated with the clinical outcome after Bonferroni adjustment for multiple hypothesis testing to account for the multiple algorithm submissions submitted (i.e., as the number of submitted results increases it is possible that some will have significant correlations in the test dataset due to chance alone).

The algorithms used different cell types for the final predictions. For example, FloReMi.1 identified a CD3+ CD4- CD45RO- CD27- CCR7- population (likely effector CD8+ cells [18], with mixed expression of CD57+ and CD57- cells. Higher frequencies of these cells early in HIV disease were correlated with worse prognosis, consistent with previous reports in chronic infection[19]. The flowType pipeline identified a CD27- TNF- CD154- population with predictive value, but this could not be interpreted further because there was either mixed or negative expression of all the other markers studied. This is not necessarily a limitation of the algorithm; it suggests that more markers were needed in the original experiment to positively identify a disease correlate. Still, the result is consistent with that obtained by FloReMi, in that CD27- cells were also identified.

Importantly, several algorithms identified populations within the CD14/VIVID+ channel, originally designed to exclude dead cells and monocytes. Cluster #3 of our independent analysis (Supplemental Figure 1) further describes this population, which includes CD3+ CD4- CD14/VIVID+ CD57- cells. This cluster was correlated with outcome both in the training and test sets. To validate this population, we constructed a gating strategy for this cell type (Supplemental Figure 3) and applied it to the unstimulated set of samples. The correlation with the clinical outcome was preserved (Supplemental Figure 4), arguing against the possibility of experimental artifact, since the identified population appeared significant in all samples (stimulated and unstimulated). Stated differently, since the population lacked cytokine-expressing cells, we expected to observe it (and find the same clinical correlations) in the unstimulated samples. Finally, we confirmed these results in a different dataset from the same cohort[10], using manual gates to identify the CD3+ CD4- CD14/VIVID+ CD57- cells. The boundaries of the gates were manually adjusted using FlowJo (Treestar Inc, Ashland, OR) to account for technical variations across the two datasets. The frequency of the manually identified cell type remained correlated with the clinical outcome (Supplemental Figure 5), confirming robustness of the results in this dataset.

We performed additional experiments to better define these cells. Based on their phenotype, we surmised that the correlation arose from dead cells or live monocytes. Notably, although the cells appeared CD3+, this was likely a consequence of well-known experimental artifacts: non-specific antibody binding by dead cells [20] and/or monocyte binding of our mouse anti-human IgG1 CD3 antibody via FCγRIIA (CD32A, [21]). To disentangle the staining patterns of the viability marker between dead cells and monocytes, we stained cells with VIVID and then measured CD14 expression using a different fluorochrome. Live, CD14+ monocytes bound intermediate levels of VIVID compared to live lymphocytes and dead monocytes (Supplemental Figure 6); a likely result of the increased size and protein content of live monocytes compared to lymphocytes[20]. This suggested that the correlation found in the dataset represented live monocytes, not dead cells (further confirmed by a separate experiment, where the correlate population expressed other markers of monocytes, CD16 and CD11c). This was confirmed by testing viability data from the original dataset, obtained using microscopic examination of ethidium bromide/acridine orange stained cells (EB/AO) at the time of thaw. There was no relationship between the frequency of dead cells, or the number of cells recovered from the vial as measured by a hemocytometer, and survival times (Supplemental Figure 7). This strongly suggests that monocytes accounted for the correlation between CD14+/VIVID+ cells in early infection and survival time, consistent with previous data [22]-[25]. This analysis demonstrates how careful analysis of biology, experimental artifacts, and computational results can provide an unbiased analysis of all available cells, revealing unexpected cell types that correlate with clinical outcomes.

## Discussion

As high throughput and high dimensional flow and mass cytometry datasets become increasingly common, there is a critical need for automated data analysis approaches. As

demonstrated by our HIV natural history studies, correlates of survival often lie within complex populations of cells not typically queried by directed, manual data analysis. Comprehensive analysis of a dataset, in which cells expressing all possible combinations of markers are examined, is unfeasible with classical, manual data analysis alone. Fortunately, there are a wide variety of approaches to automated data analysis. However, there have been few carefully controlled comparative studies that have assessed their performance. FlowCAP provides a forum for such comparison.

FlowCAP-IV was designed to address the limitations in some of the past challenges by using a dataset containing complex correlates of an externally defined variable (time to AIDS progression/diagnosis). Participants were provided a training set and their submitted results (survival status) on a blinded and independent test set. There were a number of reasons this dataset was selected for a FlowCAP challenge. It contained well-defined clinical data, providing a standardized and meaningful outcome variable. The data was also robustly collected using stringent quality control approaches, with previous studies performed in the same cohort of individuals. A variety of parameters were examined in this study, including multiple markers of T-cell maturity and function, providing a complex test. Our previous analysis also suggested that correlates of survival were contained within complex cell populations, and thus would be challenging to identify. Finally, the size of the dataset was consistent with the types of studies increasingly performed today, for which better data analysis solutions are needed.

A key product of our work with automated data analysis is the recognition that correlates of biologically important outcomes can lie in cell types that are not well described in the literature. In fact, the reliance on classically defined cell types for manual data analysis is problematic in high-dimensional experiments, as the definition of a particular cell type might differ by investigator or paper. For example, a central memory T-cell in one study may be CD45RA-CCR7+, while in another study the cell is defined as CD45RA- CD27+. When all of these markers are measured together in a sample, it becomes difficult to decide which combination of markers should define the cell type of interest and only a few cell populations are assessed due to the burden of manual analysis. In this regard, automated unbiased approaches that consider expression of all markers are vastly preferable since no potentially important cell populations are missed; this is a feature of all the automated approaches tested in this challenge. Notably, we show here that pre-gating data (to remove dead cells or monocytes, for example) can cause similar problems, as the excluded cell types may include correlates. However, this consideration is balanced by the problem of interpreting biology from the potentially artifactual staining patterns on the excluded cells.

The methods included in this study can be broken into several components: 1) Preprocessing (e.g., transformation [26] and normalization [27]); 2) feature extraction (e.g., clustering [28], probability binning [29], combinatorial gating [7], [10], and cluster matching [30], [31]); 3) supervised analysis (e.g., single-variate [32]-[34] and multi-variate models [35]); 4) characterization [8], [36] and visualization [16], [37], [38]. Not only do these components vary in their individual performance across different biological applications, their interactions with each other in a large data analysis pipeline also further complicates the choice of appropriate

methods [39] The establishment of objective benchmarks like the one reported here enables further analysis in which different components are examined subject to a wide range of technical and biological variations to identify the combination with optimal performance. It is also possible that the optimal combination may differ by data type or study design. FlowCAP provides a platform for future evaluation of these possibilities.

## Availability

Raw data, clinical meta-data, and the submissions by the participants including software are publically available at <http://flowrepository.org>[40], under repository ID FR-FCM-ZZ99 and (Note: this code will be activated only upon publication. Reviewer access is available through: <http://flowrepository.org/id/RvFrXVcrR4m0jEcsoPpnEdziUfaVUtGYdXZw4bEqvlnGfRX9aSYNSltB4bdTG4DH>).

## Acknowledgments

This work was supported by the International Society for Advancement of Cytometry's Scholarship (NA and PKC), an Ann Schreiber Mentored Investigator Award from the Ovarian Cancer Research Fund OCRF 292495 (NA), a Canadian Institute of Health Research Postdoctoral Fellowship 321510 (NA), a Canadian Institute of Health Research Scholarship for Strategic Training in Bioinformatics (NA), University of British Columbia's 4YF Scholarship (NA), Rachford and Carlota A. Harris Professorship (GPN), and the following grants: NIH 152175.5041015.0412, 1R33CA183692-01 U19 AI057229, U54CA149145, N01-HV-00242, 1U19AI100627, 5R01AI07372405, 1R33CA18365401, 1R33CA183692, R01CA184968, 1 R33 CA183654, A1073724, CA184968, and R33 CA183692 (GPN), DOD W81XWH-12-1-0591 (GPN), William Lawrence & Blanche Hughes Foundation (GPN), DOD (Department of Defense) UCD - W81XWH-13-BCRP-BREAKTHROUGH (GPN), Ovarian Cancer Teal Innovator Award (GPN), Entertainment Industry Foundation (NWCRA) (GPN), Lymphoma Research Foundation (GPN), Bill and Melinda Gates Fdn. OPP 1017093 GF12421137101 (GPN), Alliance for Lupus Research (GPN), Northrop-Grumman Corp 7500108142, NIH/NIBIB EB008400 (RRB, MM, NA, GF, RG, RS), NSERC (RB, MM), NIH/NIAID R24AI054953 (TM) NSERC (RB, MM), Agency for Innovation by Science and Technology Ph.D. grant (SVG), Research Foundation – Flanders Postdoctoral Fellowship (CVNCI R01CA163481 (PQ) Polish National Science Centre grant 2011/01/N/ST6/07035 (MBK). The authors wish to acknowledge Canada's Michael Smith Genome Sciences Centre, Vancouver, Canada for use of High Performance Computing resources.

- [1] K. O'Neill, N. Aghaeepour, J. Spidlen, and R. Brinkman, "Flow cytometry bioinformatics.," *PLoS Comput. Biol.*, vol. 9, no. 12, p. e1003365, Dec. 2013.
- [2] N. Aghaeepour, G. Finak, FlowCAP Consortium, DREAM Consortium, H. Hoos, T. R. Mosmann, R. Brinkman, R. Gottardo, and R. H. Scheuermann, "Critical assessment of automated flow cytometry data analysis techniques.," *Nat. Methods*, vol. 10, no. 3, pp. 228–238, Mar. 2013.
- [3] A. C. Weintrob, A. M. Fieberg, B. K. Agan, A. Ganesan, N. F. Crum-Cianflone, V. C.

- Marconi, M. Roediger, S. L. Fraser, S. A. Wegner, and G. W. Wortmann, "Increasing Age at HIV Seroconversion From 18 to 40 Years Is Associated With Favorable Virologic and Immunologic Responses to HAART," *JAIDS Journal of Acquired Immune Deficiency Syndromes*, vol. 49, no. 1, pp. 40–47, Sep. 2008.
- [4] A. Ganesan, P. K. Chattopadhyay, T. M. Brodie, J. Qin, W. Gu, J. R. Mascola, N. L. Michael, D. A. Follmann, M. Roederer, Infectious Disease Clinical Research Program HIV Working Group, "Immunologic and virologic events in early HIV infection predict subsequent rate of progression.," *J. Infect. Dis.*, vol. 201, no. 2, pp. 272–284, Jan. 2010.
- [5] J. MacQueen, "Some methods for classification and analysis of multivariate observations," 1967, 14 ed., vol. 1, pp. 281–297.
- [6] W. A. Moore and D. R. Parks, "Update for the logicle data scale including operational code implementations.," *Cytometry A*, vol. 81, no. 4, pp. 273–277, Apr. 2012.
- [7] K. O'Neill, A. Jalali, N. Aghaeepour, H. Hoos, and R. R. Brinkman, "Enhanced flowType/RchyOptimyx: a Bioconductor pipeline for discovery in high-dimensional cytometry data.," *Bioinformatics*, vol. 30, no. 9, pp. 1329–1330, Jan. 2014.
- [8] N. Aghaeepour, A. Jalali, K. O'Neill, P. K. Chattopadhyay, M. Roederer, H. H. Hoos, and R. R. Brinkman, "RchyOptimyx: cellular hierarchy optimization for flow cytometry.," *Cytometry A*, vol. 81, no. 12, pp. 1022–1030, Dec. 2012.
- [9] M. Malek, M. J. Taghiyar, L. Chong, G. Finak, R. Gottardo, and R. R. Brinkman, "flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification.," *Bioinformatics*, Oct. 2014.
- [10] N. Aghaeepour, P. K. Chattopadhyay, A. Ganesan, K. O'Neill, H. Zare, A. Jalali, H. H. Hoos, M. Roederer, and R. R. Brinkman, "Early immunologic correlates of HIV protection can be identified from computational analysis of complex multivariate T-cell flow cytometry assays.," *Bioinformatics*, vol. 28, no. 7, pp. 1009–1016, Apr. 2012.
- [11] D. L. Tong, G. R. Ball, and A. G. Pockley, "gEM/GANN: A multivariate computational strategy for auto-characterizing relationships between cellular and clinical phenotypes and predicting disease progression time using high-dimensional flow cytometry data," *Cytometry Part A*, 2015.
- [12] S. Van Gassen, B. Callebaut, and M. J. Van Helden, "FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data," *Cytometry Part ...*, 2015.
- [13] D. L. Tong and R. Mintram, "Genetic Algorithm-Neural Network (GANN): a study of neural network activation functions and depth of genetic algorithm search applied to feature selection," *International Journal of Machine Learning and Cybernetics*, vol. 1, no. 1, pp. 75–87, 2010.
- [14] D. L. Tong and A. C. Schierz, "Hybrid genetic algorithm-neural network: feature extraction for unprocessed microarray data.," *Artif Intell Med*, vol. 53, no. 1, pp. 47–56, Sep. 2011.
- [15] W. T. Rogers and H. A. Holyst, "FlowFP: A Bioconductor Package for Fingerprinting Flow Cytometric Data.," *Adv Bioinformatics*, vol. 2009, no. 7, pp. 193947–11, 2009.
- [16] P. Qiu, E. F. Simonds, S. C. Bendall, K. D. Gibbs Jr, R. V. Bruggner, M. D. Linderman, K. Sachs, G. P. Nolan, and S. K. Plevritis, "Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE," *Nature Biotechnology*, vol. 29, no. 10, pp. 886–891, Oct. 2011.
- [17] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*. Vol. 382. John Wiley & Sons, 2007.
- [18] P. K. Chattopadhyay and M. Roederer, "Good cell, bad cell: Flow cytometry reveals T<sub>H</sub> cell subsets important in HIV disease," *Cytometry A*, vol. 77, no. 7, pp. 614–622, Jul. 2010.
- [19] L. Papagno, C. A. Spina, A. Marchant, M. Salio, N. Rufer, S. Little, T. Dong, G.

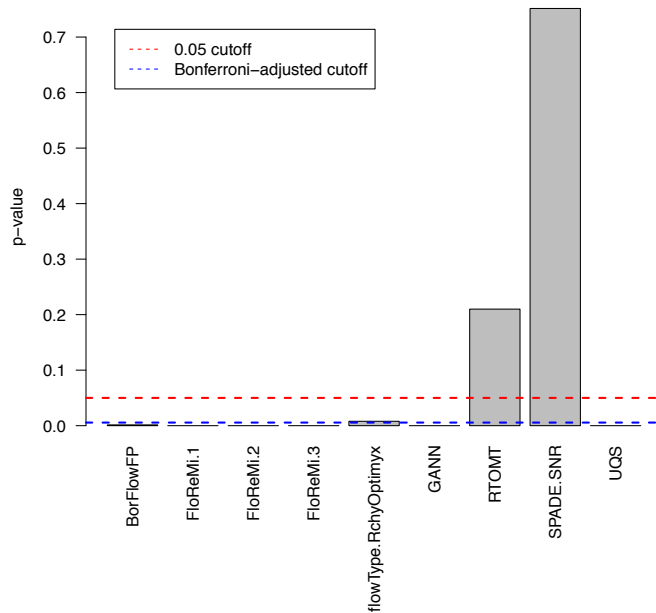


- Chesney, A. Waters, P. Easterbrook, P. R. Dunbar, D. Shepherd, V. Cerundolo, V. Emery, P. Griffiths, C. Conlon, A. J. McMichael, D. D. Richman, S. L. Rowland-Jones, and V. Appay, "Immune Activation and CD8+ T-Cell Differentiation towards Senescence in HIV-1 Infection," *PLoS Biol.*, vol. 2, no. 2, p. e20, Feb. 2004.
- [20] S. P. Perfetto, P. K. Chattopadhyay, L. Lamoreaux, R. Nguyen, D. Ambrozak, R. A. Koup, and M. Roederer, "Amine reactive dyes: An effective tool to discriminate live and dead cells in polychromatic flow cytometry," *Journal of Immunological Methods*, vol. 313, no. 1, pp. 199–208, Jun. 2006.
- [21] P. Bruhns, "Properties of mouse and human IgG receptors and their contribution to disease models," *Blood*, vol. 119, no. 24, pp. 5640–5649, Jun. 2012.
- [22] M. Müller-Trutwin and A. Hosmalin, "Role for plasmacytoid dendritic cells in anti-HIV innate immunity.," *Immunol. Cell Biol.*, vol. 83, no. 5, pp. 578–583, Oct. 2005.
- [23] R. L. Sabado, M. O'Brien, A. Subedi, L. Qin, N. Hu, E. Taylor, O. Dibben, A. Stacey, J. Fellay, K. V. Shianna, F. Siegal, M. Shodell, K. Shah, M. Larsson, J. Lifson, A. Nadas, M. Marmor, R. Hutt, D. Margolis, D. Garmon, M. Markowitz, F. Valentine, P. Borrow, and N. Bhardwaj, "Evidence of dysregulation of dendritic cells in primary HIV infection.," *Blood*, vol. 116, no. 19, pp. 3839–3852, Nov. 2010.
- [24] N. G. Sandler, H. Wand, A. Roque, M. Law, M. C. Nason, D. E. Nixon, C. Pedersen, K. Ruxrungtham, S. R. Lewin, S. Emery, J. D. Neaton, J. M. Brenchley, S. G. Deeks, I. Sereti, D. C. Douek, INSIGHT SMART Study Group, "Plasma levels of soluble CD14 independently predict mortality in HIV infection.," *J. Infect. Dis.*, vol. 203, no. 6, pp. 780–790, Mar. 2011.
- [25] A. Hasegawa, H. Liu, B. Ling, J. T. Borda, X. Alvarez, C. Sugimoto, H. Vinet-Oliphant, W.-K. Kim, K. C. Williams, R. M. Ribeiro, A. A. Lackner, R. S. Veazey, and M. J. Kuroda, "The level of monocyte turnover predicts disease progression in the macaque model of AIDS.," *Blood*, vol. 114, no. 14, pp. 2917–2925, Oct. 2009.
- [26] Y. Qian, Y. Liu, J. Campbell, E. Thomson, Y. M. Kong, and R. H. Scheuermann, "FCSTrans: An open source software system for FCS file conversion and data transformation," *Cytometry A*, vol. 81, no. 5, pp. 353–356, May 2012.
- [27] F. Hahne, A. H. Khodabakhshi, A. Bashashati, C.-J. Wong, R. D. Gascoyne, A. P. Weng, V. Seyfert-Margolis, K. Bourcier, A. Asare, T. Lumley, R. Gentleman, and R. R. Brinkman, "Per-channel basis normalization methods for flow cytometry data.," *Cytometry A*, vol. 77, no. 2, pp. 121–131, Feb. 2010.
- [28] Y. Ge and S. C. Sealfon, "flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding.," *Bioinformatics*, vol. 28, no. 15, pp. 2052–2058, Aug. 2012.
- [29] M. Roederer, A. Treister, W. Moore, and L. A. Herzenberg, "Probability binning comparison: A metric for quantitating univariate distribution differences," vol. 45, no. 1, pp. 37–46, Sep. 2001.
- [30] A. Azad, A. Khan, B. Rajwa, S. Pyne, and A. Pothen, "Classifying Immunophenotypes With Templates From Flow Cytometry," presented at the the International Conference, New York, New York, USA, 2007, pp. 256–265.
- [31] "A Non-parametric Bayesian Model with Random Effects: Joint Cell Clustering and Cluster Matching for Anomalous Sample Phenotype Identification," pp. 1–26, Feb. 2014.
- [32] V. G. Tusher, R. Tibshirani, and G. Chu, "Significance analysis of microarrays applied to the ionizing radiation response.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 9, pp. 5116–5121, Apr. 2001.
- [33] R. V. Bruggner, B. Bodenmiller, D. L. Dill, R. J. Tibshirani, and G. P. Nolan, "Automated identification of stratifying signatures in cellular subpopulations.," *PNAS*, vol. 111, no. 26, pp. E2770–7, Jul. 2014.

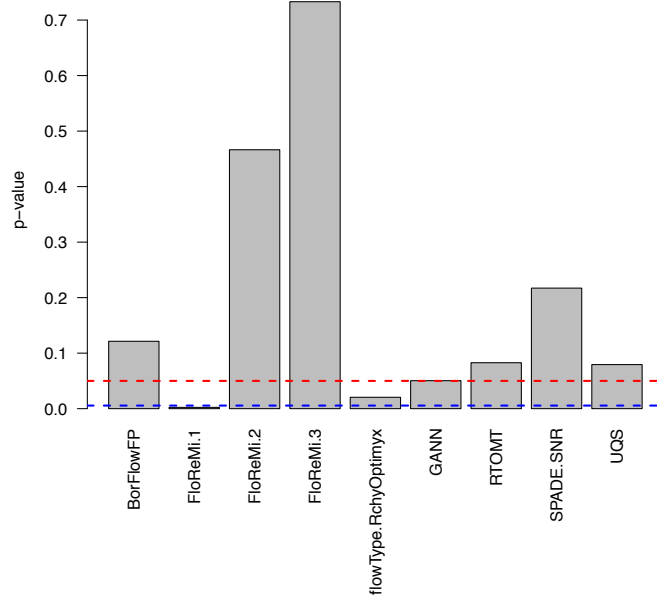
- [34] B. Gaudillière, G. K. Fragiadakis, R. V. Bruggner, M. Nicolau, R. Finck, M. Tingle, J. Silva, E. A. Ganio, C. G. Yeh, W. J. Maloney, J. I. Huddleston, S. B. Goodman, M. M. Davis, S. C. Bendall, W. J. Fantl, M. S. Angst, and G. P. Nolan, "Clinical recovery from surgery correlates with single-cell immune signatures.," *Sci Transl Med*, vol. 6, no. 255, pp. 255ra131–255ra131, Sep. 2014.
- [35] H. Zare, G. Haffari, A. Gupta, and R. R. Brinkman, "Scoring relevancy of features based on combinatorial analysis of Lasso with application to lymphoma diagnosis.," *BMC Genomics*, vol. 14, no. 1, p. S14, 2013.
- [36] B. Anchang, M. T. Do, X. Zhao, and S. K. Plevritis, "CCAST: A Model-Based Gating Strategy to Isolate Homogeneous Subpopulations in a Heterogeneous Population of Single Cells," *PLoS Comput. Biol.*, vol. 10, no. 7, p. e1003664, Jul. 2014.
- [37] E.-A. D. Amir, K. L. Davis, M. D. Tadmor, E. F. Simonds, J. H. Levine, S. C. Bendall, D. K. Shenfeld, S. Krishnaswamy, G. P. Nolan, and D. Pe'er, "viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia.," *Nature Biotechnology*, vol. 31, no. 6, pp. 545–552, Jun. 2013.
- [38] K. Shekhar, P. Brodin, M. M. Davis, and A. K. Chakraborty, "Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE).," *PNAS*, vol. 111, no. 1, pp. 202–207, Jan. 2014.
- [39] G. Finak, J. Frelinger, W. Jiang, E. W. Newell, J. Ramey, M. M. Davis, S. A. Kalams, S. C. De Rosa, and R. Gottardo, "OpenCyto: an open source infrastructure for scalable, robust, reproducible, and automated, end-to-end flow cytometry data analysis.," *PLoS Comput. Biol.*, vol. 10, no. 8, p. e1003806, Aug. 2014.
- [40] J. Spidlen, K. Breuer, C. Rosenberg, N. Kotecha, and R. R. Brinkman, "FlowRepository: A resource of annotated flow cytometry datasets associated with peer-reviewed publications," *Cytometry A*, vol. 81, no. 9, pp. 727–731, Sep. 2012.

**Table 1: A summary of all submitted results.**

Name	Approach	Availability	Reference
BorFlowFP	Exhaustive all-relevant feature selection with the Boruta method over the flow cytometry fingerprints data followed by modeling with Survival Random Forest algorithm.	Boruta and FlowFP are free, open-source R packages available from CRAN and BioConductor, respectively.	[15]
FloReMi.1	Partitioning of cells (e.g., into positive or negative) to identify cell populations, followed by feature extraction of both cell population size and MFI values. Features with minimal redundancy were selected as input for a random survival forest for survival time prediction.	R Code to reproduce our results is available at <a href="http://www.github.com/SofieVG/FloReMi">www.github.com/SofieVG/FloReMi</a> and in Supplementary data from FlowRepository.org	[12]
FloReMi.2	As FloReMi.1, but using a Cox-proportional hazards model for survival time prediction	R Code to reproduce results is available at <a href="http://www.github.com/SofieVG/FloReMi">www.github.com/SofieVG/FloReMi</a> and in supplementary data from FlowRepository.org	[12]
FloReMi.3	As FloReMi.1, but using an additive hazards model for survival time prediction	R Code to reproduce results is available at <a href="http://www.github.com/SofieVG/FloReMi">www.github.com/SofieVG/FloReMi</a> and in supplementary data from FlowRepository.org	[12]
flowDensity/flowType/ RchyOptimyx	Density-based getting and partitioning of cells (e.g., into positive or negative) followed by dynamic programming to ID k-shortest paths to important cell populations	Free, open-source R packages available from BioConductor	[10,7,8,9]
GANN	Identifying profiles of individual bin channels of fluorescence for the different cell markers which can be used for distinguishing different groups / categories	C++ written program, available by contacting the author and in supplementary data from FlowRepository.org	[11,13,14]
RTOMT	Partitioning of cells (e.g., into positive or negative). Combining clinical diagnosis and survival time into a single target vector. Regression tree.	Free, open-source code, Matlab available in supplementary data from FlowRepository.org	NA
SPADE.SNR	SPADE was used to derive cell clusters in high-dimensional space. Important clusters were selected by a signal-to-noise ratio based on cell frequency of subjects who showed disease progression.	Freely available implementation in Matlab	16
EMMIXflow	Clustering using expectation maximization fitting of skew-t mixture models was used for feature extraction. A random survival forest was used for building the predictive model.	R source code available in supplementary data from FlowRepository.org	17



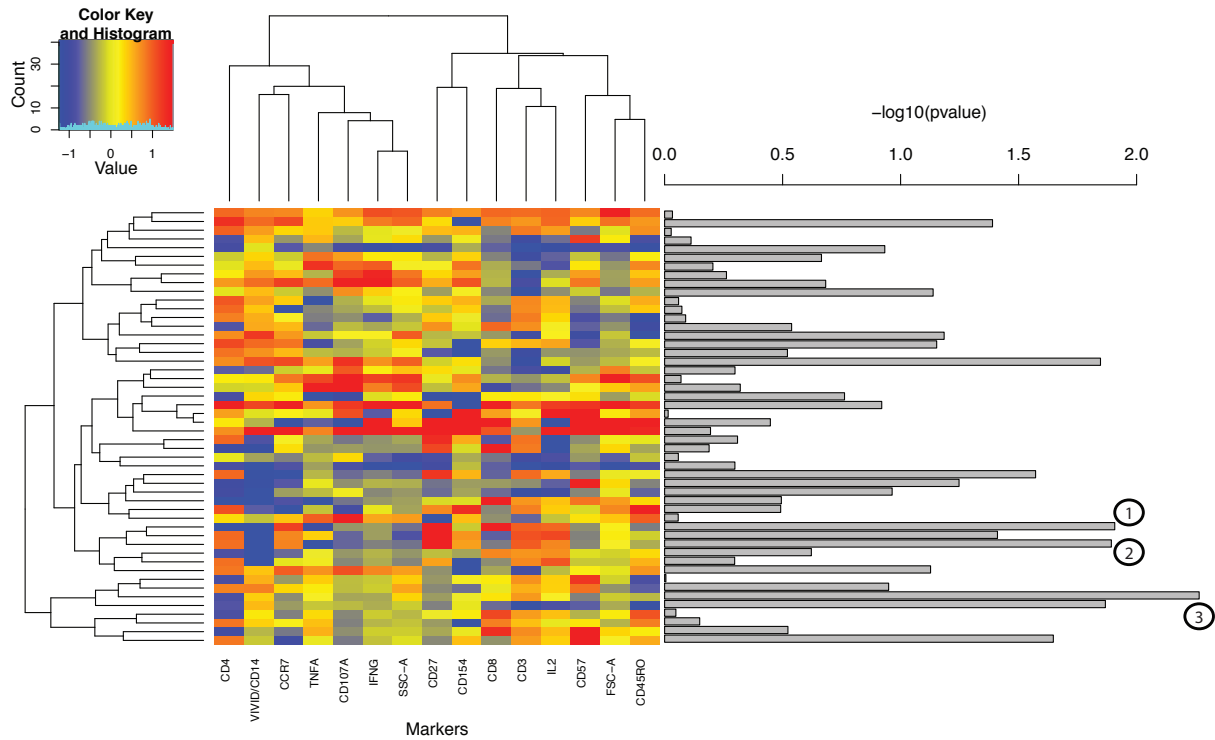
(A)



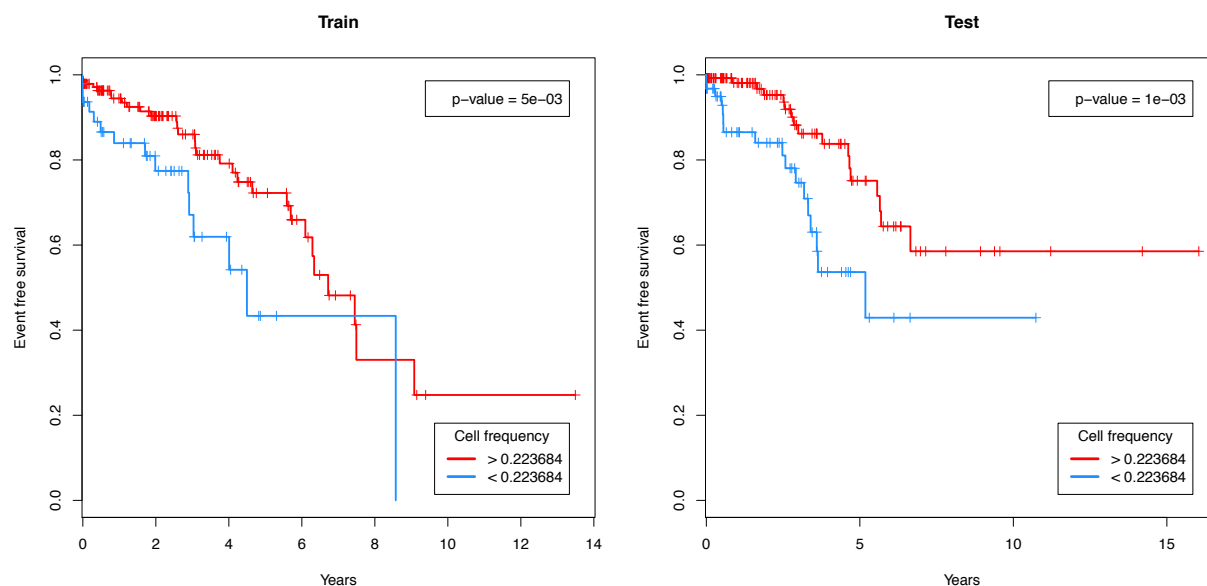
(B)

Figure 1: The log-rank test p-values of the submissions on (A) the training-set and (B) the test-set.

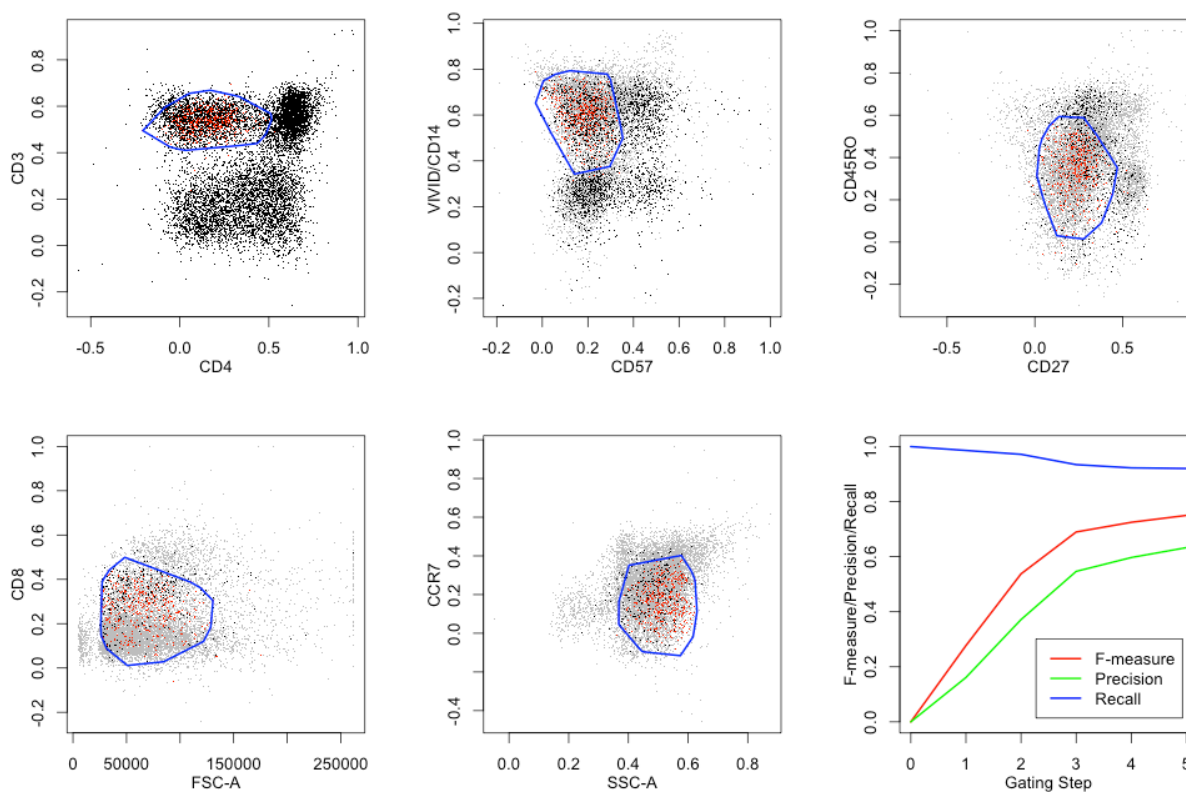
## Supplemental Information



**Supplemental Figure 1: Independent analysis of the FlowCAP-IV dataset. 50 cell-types were identified using the K-means clustering algorithm. The heatmap demonstrates the median expression for each marker (column) in the cells of each cluster (rows). The p-values of the log-rank tests on the training set are visualized using bar plots on the right.**

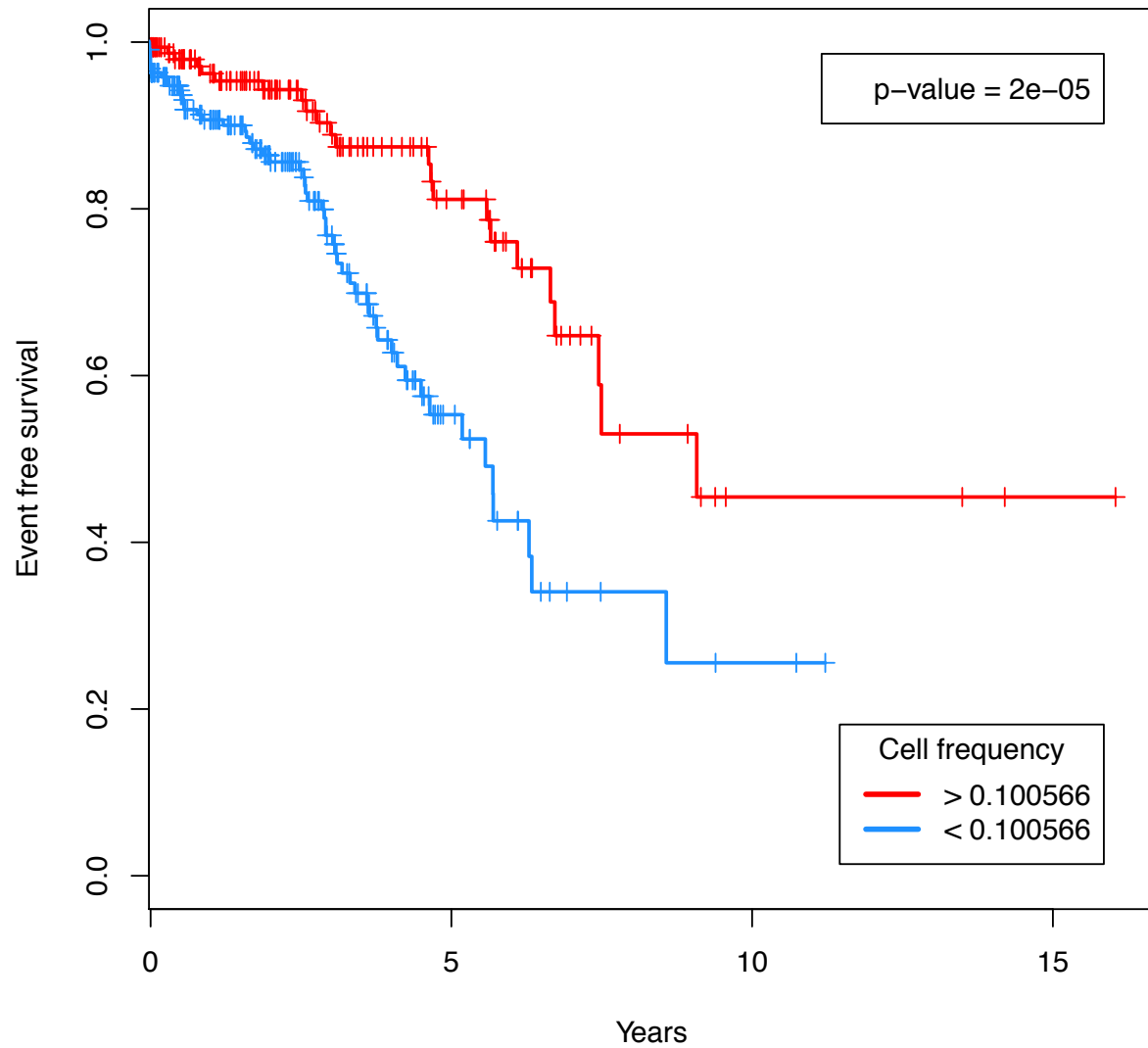


**Supplemental Figure 2: Kaplan-Meier curves for the best cell type (#3) from the independent K-means analysis on the training and test dataset. The patients were divided into two groups using the best possible threshold by testing all percentiles.**

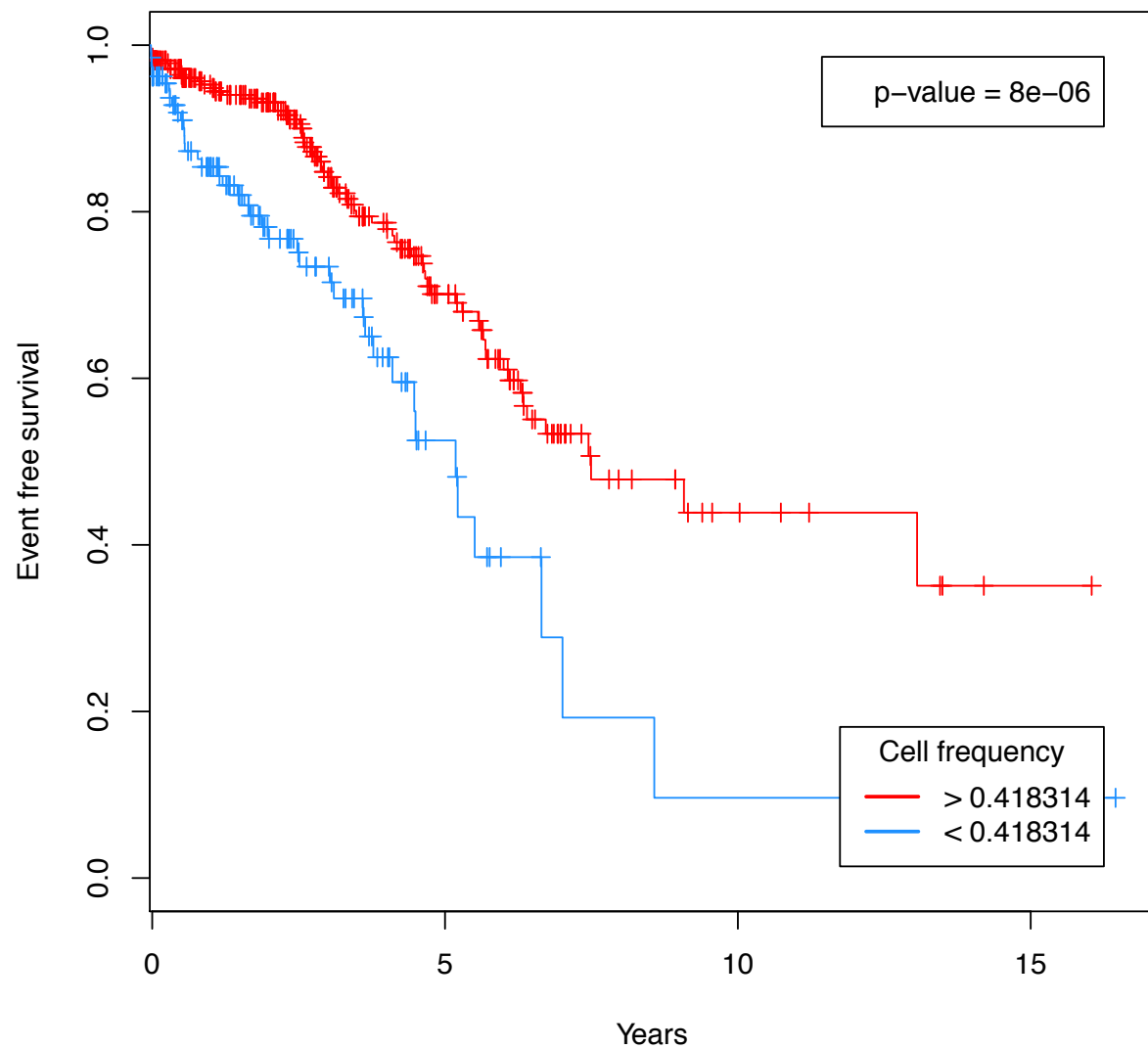


**Supplemental Figure 3: A manual gating strategy for identification of the cluster most correlated with the clinical outcome from the independent K-means analysis (cluster #3, Figure 1). The polygons (starting from top-left moving to right) can recapitulate the high-dimensional cell type with the purity, recall, and F-measure (average of precision and recall) visualized on the last panel (bottom-right). The red dots are the cells of**

interest, the black dots are all other cells, and the gray dots are the dots excluded in the previous gating step.

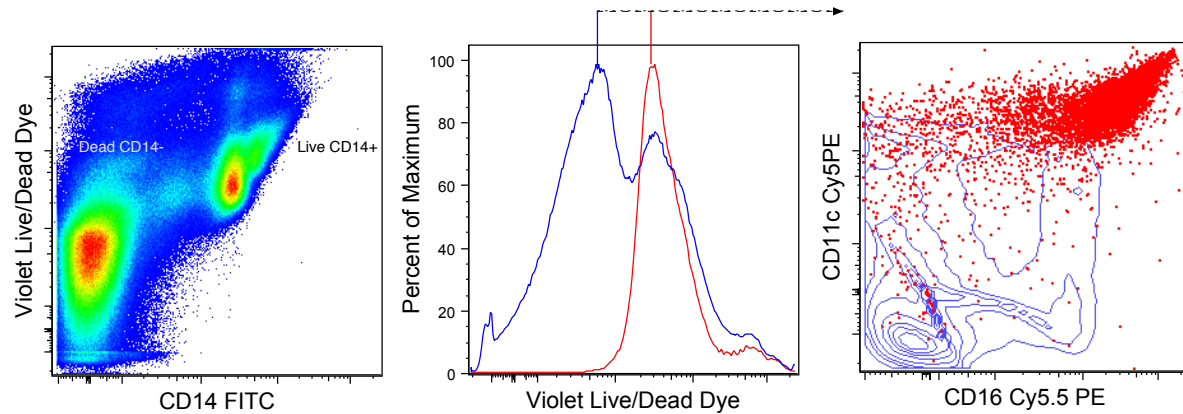


**Supplemental Figure 4: Kaplan-Meier curves for the manual gating strategy applied to the unstimulated samples to confirm robustness of the results to Gag stimulation. The patients were divided into two groups using the best possible threshold by testing all cell frequency divisions.**

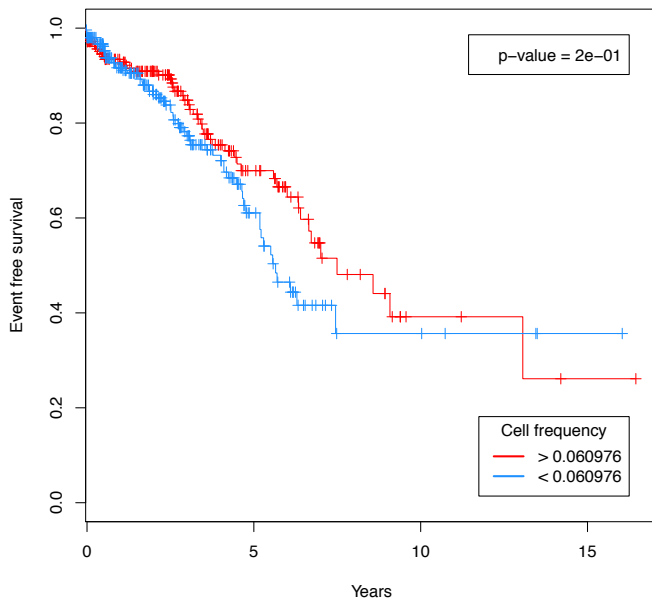


**Supplemental Figure 5: Kaplan-Meier curves for the manual gating strategy applied to a confirmatory dataset from the same cohort. The patients were divided into two groups using the best possible threshold by testing all cell frequency divisions.**

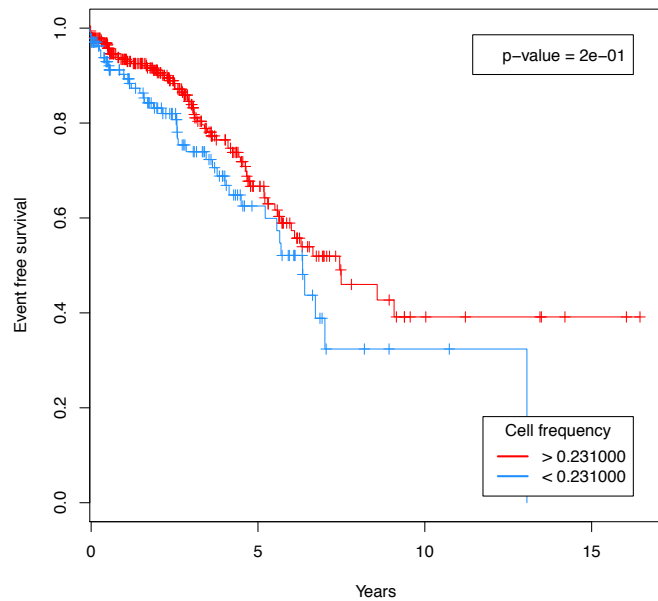




**Supplemental Figure 6: The correlate of protection found in the FlowCAP-IV dataset amongst VIVID/CD14+ cells likely represents a population of live CD14+ cells, not dead cells. When cells are stained with VIVID and CD14 using separate fluorochromes, it becomes clear that CD14+ cells bind more VIVID than CD14- cells (first panel), as a consequence of increased size and protein content. Thus, the distribution of VIVID staining for live monocytes overlaps with dead CD14- lymphocytes (second panel), and the measurement of these two markers in the same channel in the original dataset complicates biological interpretation of the data. The high expression of CD16 and CD11c by these cells (third panel) is consistent with their monocyte/myeloid classification. In the second panel the blue curve shows live/dead staining among lymphocytes, with the left peak representing live cells, and the right peak representing dead cells. The red curve shows live/dead staining in monocytes. This curve is right-shifted, such that live monocytes overlap with dead lymphocytes. Dead monocytes are found in the brightest peak of the red curve.**



**(A)**



**(B)**

**Supplemental Figure 7: Kaplan-Meier curves based on the frequency of dead cells (left) and the proportion of cells recovered from each vial (right). The patients were divided into two groups using the best possible threshold by testing all frequency divisions.**